Calculation of Mean Abundance in the Red Urchin Analysis Program and Green Urchin Analysis Program

J. Lochead, W. Hajas, and D. Leus

2015

Fisheries and Oceans Canada
Science Branch, Pacific Region
Pacific Biological Station
Nanaimo, British Columbia
V9T 6N7

Canadian Manuscript Report of Fisheries and Aquatic Sciences 3065

Fisheries and Oceans Canada | Pêches et Océans Canada

Canada

Canadian Manuscript Report of Fisheries and Aquatic Sciences

Manuscript reports contain scientific and technical information that contributes to existing knowledge but which deals with national or regional problems. Distribution is restricted to institutions or individuals located in particular regions of Canada. However, no restriction is placed on subject matter, and the series reflects the broad interests and policies of Fisheries and Oceans Canada, namely, fisheries and aquatic sciences.

Manuscript reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in the data base *Aquatic Sciences and Fisheries Abstracts*.

Manuscript reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page.

Numbers 1-900 in this series were issued as Manuscript Reports (Biological Series) of the Biological Board of Canada, and subsequent to 1937 when the name of the Board was changed by Act of Parliament, as Manuscript Reports (Biological Series) of the Fisheries Research Board of Canada. Numbers 1426 - 1550 were issued as Department of Fisheries and Environment, Fisheries and Marine Service Manuscript Reports. The current series name was changed with report number 1551.

Rapport manuscrit canadien des sciences halieutiques et aquatiques

Les rapports manuscrits contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui traitent de problèmes nationaux ou régionaux. La distribution en est limitée aux organismes et aux personnes de régions particulières du Canada. Il n'y a aucune restriction quant au sujet; de fait, la série reflète la vaste gamme des intérêts et des politiques de Pêches et Océans Canada, c'est-à-dire les sciences halieutiques et aquatiques.

Les rapports manuscrits peuvent être cités comme des publications à part entière. Le titre exact figure au-dessus du résumé de chaque rapport. Les rapports manuscrits sont résumés dans la base de données *Résumés des sciences aquatiques et halieutiques*.

Les rapports manuscrits sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page du titre.

Les numéros 1 à 900 de cette série ont été publiés à titre de Manuscrits (série biologique) de l'Office de biologie du Canada, et après le changement de la désignation de cet organisme par décret du Parlement, en 1937, ont été classés comme Manuscrits (série biologique) de l'Office des recherches sur les pêcheries du Canada. Les numéros 901 à 1425 ont été publiés à titre de Rapports manuscrits de l'Office des recherches sur les pêcheries du Canada. Les numéros 1426 à 1550 sont parus à titre de Rapports manuscrits du Service des pêches et de la mer, ministère des Pêches et de l'Environnement. Le nom actuel de la série a été établi lors de la parution du numéro 1551.

Canadian Manuscript Report of
Fisheries and Aquatic Sciences 3065

2015

CALCULATION OF MEAN ABUNDANCE IN THE RED URCHIN ANALYSIS PROGRAM AND GREEN URCHIN ANALYSIS PROGRAM

by

J. Lochead, W. Hajas, and D. Leus

Fisheries and Oceans Canada
Science Branch, Pacific Region
Pacific Biological Station
Nanaimo, British Columbia
V9T 6N7

© Her Majesty the Queen in Right of Canada, 2015

Cat. No. Fs 97-4/3065E ISBN 978-0-660-02010-5 ISSN 0706-6473 (print version)

Cat. No. Fs 97-4/3065E-PDF ISBN 978-0-660-02011-2 ISSN 1488-5387 (PDF version)

Correct citation for this publication:

Lochead, J., Hajas, W., and Leus, D. 2015. Calculation of mean abundance in the Red Urchin Analysis Program and Green Urchin Analysis Program. Can. Manuscr. Rep. Fish. Aquat. Sci. 3065: vi + 41 p.

TABLE OF CONTENTS

ABSTRACT vi

RÉSUMÉ vi

1 Introduction 1

2 Methods and Results 1

2.1 Overview 1
2.2 User Input 2
2.2.1 Transect Categories 3
2.2.2 Size Categories 3
2.2.3 Depth Range 4
2.3 Notation 4
2.3.1 Conventions 4
2.3.2 Indices 5
2.3.3 Variables 5
2.4 Estimation of Mean Abundance 6
2.4.1 Abundance for a Category of Transects 6
2.4.2 Unsurveyed Quadrats within a Transect 7
2.5 ALLOMETRIC RELATIONSHIPS 11
2.5.1 Red Urchin Allometric Relationship 11
2.5.2 Green Sea Urchin Allometric Relationship 12
2.6 Mean Weight 15
2.7 Deterministic Estimation of Biomass 16
2.8 Stochastic Estimates of Abundance 16
2.8.1 Stochastic Estimates of Number of Urchins 17
2.8.2 Stochastic Estimates of Biomass 18
2.9 The Ratio Estimator 19
2.9.1 Bootstrapping 20

2.9.2 Random Seeds 20
2.10 Outputs 21
3 DISCUSSION 21
4 ACKNOWLEDGEMENTS 22
5 REFERENCES 22
Figure 1. Linear interpolation as used to approximate urchin counts from a subsampled Red Sea Urchin transect. 24
6 Appendix A. Description of Variables 25
7 Appendix B. Software Structure 28
7.1 Appendix B.1. The UnMeasAnimals class 28
7.1.1 Input to the constructor: 28
7.1.2 Methods of Note: 28
7.1.3 Other Notes: 28
7.2 Appendix B.2. The MeasAnimals Class 29
7.2.1 Input to the constructor: 29
7.2.2 Methods of Note: 29
7.2.3 Other Notes: 29
7.3 Appendix B.3. The Quadrat Class 30
7.3.1 Input to the constructor: 30
7.3.2 Methods of Note: 30
7.4 Appendix B.4 The Transect Class 31
7.4.1 Input to the constructor: 31
7.4.2 Methods of Note: 31
7.4.3 Other Notes: 31
7.5 Appendix B.5 The TransectCategory Class 32
7.5.1 Input to the constructor: 32

7.5.2 Methods of Note: 32

7.6 Appendix B.6 The DataAllometric Class 33
7.6.1 Input to the constructor: 33

7.6.2 Methods of Note: 33

7.6.3 Other Notes: 33

7.7 Appendix B.7 The GSU Library 34
7.7.1 Data Contained in the library: 34

7.7.2 Functions to generate SQL commands: 34

7.8 Appendix B.8 The RSU Library 36
7.8.1 Data Contained in the library: 36

7.8.2 Functions to generate SQL- commands: 36

7.9 Appendix B. References 38

Table B.1. Data from the Red Sea Urchin and Green Sea Urchin biological databases used by RUAP and GUAP to estimate mean density. Key and Hkey are used to construct joins between tables within the database. QuadratNum is used to construct joins of quadrats within a transect. 39

Figure B.1. Red Sea Urchin biological database relationships 40

Figure B.2. Green Sea Urchin biological database relationships 41

ABSTRACT

Lochead, J., Hajas, W., and Leus, D. 2015. Calculation of mean abundance in the Red Urchin Analysis Program and Green Urchin Analysis Program. Can. Manuscr. Rep. Fish. Aquat. Sci. 3065: vi + 41 p.

The Red Urchin Analysis Program (RUAP) and Green Urchin Analysis Program (GUAP) are computer programs developed by the Department of Fisheries and Oceans Canada to estimate the mean density of Red and Green Sea Urchins from fishery-independent survey data. RUAP and GUAP work directly from the same file structure that is found in the relational database that houses the survey data, and they accommodate the quadrat subsampling that routinely occurs. Density is calculated in terms of population (numbers of urchins) and biomass (grams of urchins) and is reported per square meter of seafloor as well as per meter of shoreline. Calculations are performed for specific size categories. Deterministic, best estimate calculations are made, as well as stochastic estimates that are used for establishing confidence bounds. This report documents the calculations performed by RUAP and GUAP to estimate population densities and urchin biomass from fishery-independent survey data for defined geographic areas within British Columbia.

RÉSUMÉ

Lochead, J., Hajas, W., et Leus, D. 2015. Calcul de l'abondance moyenne au moyen du programme d'analyse de l'oursin rouge et du programme d'analyse de l'oursin vert. Rapp. manus. can. sci. halieut. aquat. 3065: vi + 41 p.

Le programme d'analyse de l'oursin rouge(PAOR) et le programme d'analyse de l'oursin vert (PAOV) sont des programmes informatiques mis au point par le Ministère des Pêches et des Océans du Canada afin d'estimer la densité moyenne des oursins rouges et verts à partir de données de relevés indépendants de la pêche. Le PAOR et le PAOV fonctionnent directement à partir de la même structure de fichier qu'on retrouve dans les bases de données relationnelles qui hébergent les données des relevés et ils acceptent les sous-échantillonnages de quadrats qui surviennent constamment. La densité est calculée en termes de population (nombre d'oursins) et de biomasse (grammes d'oursins) et est déclarée par mètre carré de fond d'océan et par mètre de littoral. Les calculs sont effectués pour des catégories particulières de tailles. Des calculs déterministes, de meilleure valeur estimée, sont effectués, et des estimations stochastiques, qui servent à établir les seuils de confiance. Le présent rapport contient les calculs effectués au moyen du PAOR et du PAOV afin d'estimer les densités de populations et la biomasse des oursins à partir de données de relevés indépendants de la pêche pour des zones géographiques définies en Colombie-Britannique.

1 INTRODUCTION

The Red Urchin Analysis Program (RUAP) and Green Urchin Analysis Program (GUAP) are computer programs that were developed to support stock assessment activities on the west coast of Canada. The Red Sea Urchin (*Mesocentrotus franciscanus*) and Green Sea Urchin (*Strongylocentrotus droebachiensis*) stocks in British Columbia (BC) are assessed every three years through the analysis of fishery-dependent and fishery-independent data (DFO 2014, Leus et al. 2014). Fishery-independent data are obtained from dive surveys carried out collaboratively by the Fisheries and Oceans Canada (DFO), the Pacific Urchin Harvesters Association and the West Coast Green Urchin Association.

The dive survey protocols, which are similar for both species, were developed to survey stocks in specific areas of the BC coast (DFO 2014; Leus et al. 2014). Transect locations are either randomly selected before a survey or, for index sites, the same transects are re-surveyed. Transects run perpendicular to shore and/or depth contours, from deep water (50 feet gauge depth for Red Sea Urchins and 33 feet below chart datum for Green Sea Urchins) to shallow water (0 feet gauge depth for Red Sea Urchins and 0 feet below chart datum for Green Sea Urchins). Survey methods are as follows: divers count and measure urchins occurring within a 1 m^2 quadrat that is moved over the substrate for the length of each transect, following either a leadline or compass bearing from deep to shallow. Due to diving time limitations, the protocols include a systematic method of subsampling quadrats for improved efficiency and reduced costs. Survey protocols are documented in detail in Leus et al. 2014 for Red Sea Urchins and in DFO 2014` for Green Sea Urchins.

The current paper describes how the calculations within RUAP and GUAP estimate urchin biomass and population densities from fishery-independent survey data for defined geographic areas.

2 METHODS AND RESULTS

2.1 OVERVIEW

The goal of GUAP and RUAP is to estimate urchin density and biomass, spatially and linearly, from dive survey data. For the purpose of density and biomass calculations, the sampling units are transects.

The programs were designed to accommodate survey protocols which intentionally collect partial datasets due to time limitations associated with SCUBA diving. Typically, every second quadrat is surveyed while every other is skipped (e.g. sample odd numbered quadrats and skip even numbered quadrats). For surveyed quadrats, either urchin count is recorded or urchin count and test diameters of urchins in the quadrat. The pattern of counts and measurements for

surveyed quadrats varies from transect to transect. For some transects, measurements are taken in all surveyed quadrats, while on other transects measurements are taken on every second or third surveyed quadrat, and counts-only are recoded for other surveyed quadrats. RUAP and GUAP use linear interpolation to fill in the gaps, based on the assumption that population density, depth and proportion of urchins in each size category change linearly between surveyed quadrats of the same transect. Allometric relationships are used to convert test diameters to estimated weights, which in turn are used to estimate biomass.

Density is calculated in terms of population (numbers of urchins) and biomass (grams of urchins) and is reported spatially (per square metre of seafloor) as well as linearly (per metre of shoreline). The results of RUAP and GUAP include confidence bounds on the estimated densities of Red and Green Sea Urchins. The confidence bounds are provided to fishery managers as part of the science advice pertaining to these wild Red and Green Sea Urchin populations (DFO 2014; Leus et al. 2014).

The process of estimating densities involves, interpolation of depth and of number of urchins for unsurveyed quadrats, interpolation of the proportion of urchins in each size category for un-measured quadrats (unsurveyed and count-only surveyed quadrats), application of depth restrictions (if specified), calculation of density through a ratio estimator and bootstrapping to estimate confidence bounds for the category of transects chosen. The process of estimating biomass is the same with the addition of calculations of mean weight of urchins in each size category on the transect (from size frequency data and length-weight allometric relationships).

Two methods are used for interpolation. First, deterministic interpolation is used to derive density and biomass estimates, which are also used as inputs in the BCA bootstrapping process (see Section 2.9.1). Second, stochastic interpolation is used in creating transect estimates used in re-sampling of transect values during the bootstrapping process.

The following sections provide mathematical descriptions of the calculations incorporated within RUAP and GUAP. The software used to implement GUAP and RUAP is described in Appendix B.

2.2 USER INPUT

Data is read from a Microsoft SQL Server database accessed via Microsoft Access (.mdb or .accdb) employing Open Database Connectivity (ODBC) using the OLE DB version 12 (Microsoft) library. RUAP and GUAP are compatible with the species specific views generated from the biological database maintained by the DFO Shellfish Data Unit (Figures B.1 and B.2), or any other MS Access file (versions 97 to 2010) with the data structure shown in Table B.1The user specifies the input database file to use.

RUAP identifies all combinations of *Location* and *Year*, in the input database file. From these combinations, the user selects the surveys to include in the analyses. RUAP then uses all the data from the selected survey(s) from the corresponding year(s).

GUAP allows the user to select data according to combinations of *SurveyTitle* and *Year*. Therefore specific surveys can be run for particular years.

The programs also allow the user to change the number of iterations for the bootstrap process (Section 2.9.1) or the random seed (Section 2.9.2).

2.2.1 Transect Categories

Once the surveys are selected, the user selects the properties used to define categories of transects. The properties correspond to fields in the Headers table of the input databases (Table B.1). Typically, users will require density estimates for specific Pacific Fisheries Management Areas or SubAreas. The user can select "StatArea" and/or "SubArea" from the GUAP or RUAP interface to define transect categories according to these properties. Urchin density is then calculated separately for each PFM Area or Subarea.

In RUAP, transect characteristics that the user can select from are: SurveyTitle, Location, SiteNum, Year, StatArea, SubArea, InBed.

In GUAP, transect characteristics that the user can select from are: Survey title, Sub-sample location, Year, Statistical area and Statistical subarea, or combinations of those.

2.2.2 Size Categories

Within each transect category, mean density is estimated for specific size categories. Size is measured according to the test diameter (TD) of the urchins as it is recorded in the database (Table B.1). It is assumed each urchin has a test diameter between zero and one thousand millimetres.

With RUAP, the user may choose to divide the size range into subintervals to create multiple size categories. The upper size limits for these subintervals are restricted to 50, 70, 89, 110, 140 and 1000 mm. The legal minimum size for the Red Sea Urchin commercial fishery is 90 mm. So, for example, if 89 and 140 mm are used as subinterval maxima, then the resulting size categories corresponds to size categories of 0 to 89, 90 to 140 and 141 to 1000 millimetres. The size categories typically used in RUAP are 0 to 89 and 90 to 1000 millimetres, i.e. sublegal and legal sized urchins.

Since green urchins are generally smaller than red urchins, different size categories are made available in GUAP. The upper size limits for the green urchin subintervals are restricted to 24, 39, 54, 69, 84 and 1000 mm. The legal minimum size for the Green Sea Urchin commercial fishery is 55 mm. Calculations are typically performed for urchins that are legal-sized (55 mm and larger), sublegal-mature (25mm $\leq$ TD $<$ 55 mm), and sublegal-immature (24 mm and smaller).

Mean abundance is simultaneously estimated for all size categories. In any single application of RUAP or GUAP the size categories do not overlap. Overlapping size categories can be achieved by running RUAP or GUAP multiple times.

2.2.3 Depth Range

Users can also specify a minimum and maximum depth to be considered in the analysis. For the green urchin program, only data collected between 0 and 10 m below chart datum are used for quota recommendations and survey reports (Waddell and Perry 2012). For the red urchin program, depth truncation is not used for quota recommendations; however density by depth range is often computed for survey reports.

2.3 NOTATION

A description of variables and indices used by the programs are provided in Appendix A.

2.3.1 Conventions

Lowercase letters are used to represent quadrat specific values while uppercase letters are used to represent transect specific values. For example, b represents the biomass in a quadrat while B represents the biomass for an entire transect. Greek letters are used to represent values for individual urchins.

The prefixes ^{m}, ^{u} and ^{t} are used to distinguish between measured, unmeasured and total urchins. For example, $^{m}b_{T,q}$, $^{u}b_{T,q}$ and $^{t}b_{T,q}$ represent the biomass of measured, unmeasured and all urchins in transect T and quadrat q.

A prefix of * is used to indicate randomly sampled values or values calculated using randomly sampled values. The indices are defined in section 2.3.2.

An arrow, $\rightarrow$, is used to indicate an array of size category-specific values. For example, ${}^{m}\vec{u}_{T,q} = \left({}^{m}u_{T,q,1}, {}^{m}u_{T,q,2}, \ldots, {}^{m}u_{T,q,F}\right)$ represents the number of measured urchins in each size category (1 to F) in quadrat q on transect T.

2.3.2 Indices

- T is an index for transects.
- q is an index for the relative position of quadrats within a transect. $q = 1,2,\ldots,Q_T$ and includes both surveyed and unsurveyed quadrats.
- $f = 1,2,\ldots,F$ is the index for size category; typically only 2 or 3 size categories are used.
- $r = 1,2,\ldots,{}^{m}u_{T,q}$ is the index for measured urchins in a quadrat.
- $i = 1,2,\ldots,$ represents iterations in a bootstrapping process.

2.3.3 Variables

Q_T is the number of quadrats in a transect, in the specified depth range. It includes surveyed quadrats (for which there are survey data) and unsurveyed quadrats (for which there is no survey data, i.e., quadrats skipped by divers where data is interpolated).

U and u represent the number of urchins. U only includes urchins from quadrats that are within the specified depth range while u represents the number of urchins in a given quadrat.

B and b represent the biomass, or weight, of urchins. B only includes biomass from quadrats that are within the specified depth range while b represents biomass in a given quadrat.

$\vec{p}_{T,q}$ is the probability associated with each size category for quadrat q on transect T:
$\sum \vec{p}_{T,q} = 1$

$\vec{W}_T$ is the estimated mean weight of urchins (grams) for each size category, f, on transect T.

$A_{T,f}$ is the abundance of urchins in a transect, T, in a given size category, f, that occur within the specified depth range. A_T is generally expressed in terms of the number of urchins (population) or the weight of urchins (biomass).

S_T is the size of a transect. For linear calculations transect-size corresponds to the width of the transect, i.e., one meter. For spatial calculations, transect-size corresponds to the area of the transect estimated to be within the specified depth range and is equal to the number of quadrats in the specified depth range.

D_f is an estimate of mean density for a group of transects, in a given size category, f, and is calculated in terms of population (numbers of urchins) and biomass (grams of urchins).

$\Omega_{T,q,r}$ is the test diameter (mm) of an urchin r in quadrat q on transect T.

$\Psi_{T,q,r}$ is the estimated weight (grams) of an urchin r in quadrat q on transect T.

$d_{T,q}$ is the depth (metres) of the quadrat q of transect T.

2.4 ESTIMATION OF MEAN ABUNDANCE

2.4.1 Abundance for a Category of Transects

Abundance is measured in two ways:

1. Population (number of urchins)
2. Biomass (grams of urchins)

A category of transects will generally represent a sample from a specific location. For the purpose of managing an urchin fishery, the size of a location can also be measured in two ways:

a. Linearly (metres of shoreline)
b. Spatially (square metres of sea floor)

As a result, there are four different ways that mean abundance is reported for each combination of transect and size categories chosen by the user:

1a. Linear Population
1b. Spatial Population
2a. Linear Biomass
2b. Spatial Biomass

2.4.2 Unsurveyed Quadrats within a Transect

Ideally, a transect would be fully surveyed with the depth of every quadrat recorded and every urchin counted and measured. However, quadrats within a transect are often subsampled. In some cases, the subsampling is a strategic part of the survey plan that allows the scuba divers to spend less time per transect and survey more transects. For example, the survey protocol may only require that urchins from every second quadrat be counted and urchins from every fourth quadrat be measured. In other cases, the decision to subsample quadrats is made by divers in the water in order address immediate logistical and safety concerns.

The vast majority of quadrats fall into one of three categories:

1. All the urchins are counted and measured and depth is recorded. (Surveyed measured)
2. All the urchins are counted and depth is recorded but no urchins are measured. (Surveyed counted)
3. No measurements or counts or depths are recorded. (Unsurveyed)

In order to determine if quadrats are within the targeted depth range, a suitable approximation is required for the quadrats where depth was not recorded. From an analysis perspective, the unsurveyed quadrats can occur on an arbitrary basis. The surveyed quadrats do not constitute a random sample and they do not necessarily occur at regular intervals along the transect.

When approximating data to unsurveyed quadrats, it is important to note that the transects can be long. Over half the urchin transects with unsurveyed quadrats are longer than forty metres. There is no reliable representation of the characteristics of the transect as a whole (hence the need to interpolate data to the unsurveyed quadrats). Therefore, nearby quadrats are the preferred basis for interpolating to unsurveyed quadrats.

It should also be noted that RUAP and GUAP are intended to perform calculations on a large scale and on an automated basis; without the opportunity to respond to the quirks of individual transects. A very robust methodology is required to approximate data to unsurveyed quadrats.

Linear interpolation is an approach that is robust and that reflects the assumption that quadrats in close proximity are similar to each other (spatial autocorrelation).

2.4.2.1 Depth

RUAP and GUAP take quadrat depths from the corresponding biological database (see details in Appendices B.3, B.8 and B.9). When quadrat depth is unavailable from the data, it is estimated by linear interpolation using the preceding and proceeding quadrats' depth.

For each transect, there is assumed to be at least one quadrat where depth, $d_{T,q}$, has been recorded as part of the survey data. Where m is the first quadrat where depth has been recorded, any preceding quadrats that are missing depth data are assigned the same depth as m:

$$d_{T,q} = d_{T,m} \text{ for } q < m$$

Similarly, where n is the last quadrat where depth has been recorded, any quadrats lacking depth data that follow are assigned the same depth as n:

$$d_{T,q} = d_{T,n} \text{ for } q > n$$

Missing depth data almost never occurs at the very beginning or end of a transect. This functionality has been built-in to ensure the program could deal with this, in the event that it arises in the future.

For a quadrat, q, that occurs between quadrats j and k that have recorded depth, linear interpolation is used to estimate the depth:

$$d_{T,q} = d_{T,j} + \frac{q-j}{k-j} * \left(d_{T,k} - d_{T,j}\right)$$

The user has the option to restrict density and biomass calculations to only include quadrats from a specific depth range of recorded or estimated depths. Such a restriction can reduce the number of urchins in a transect if quadrats (and their associated urchins) occur outside the selected depth range. For spatial calculations of density, a depth restriction can also reduce the size of a transect, as each quadrat excluded removes 1 m^2 of sea floor. For linear calculations of density, transect size is not affected by depth restrictions, as each transect remains 1 m wide regardless of the number of quadrats incorporated into the calculations.

2.4.2.2 Deterministic Approximation of Number of Urchins in a Quadrat

For each transect, there is assumed to be at least one "counted" quadrat where total count, ${}^{t}u_{T,q}$, has been recorded in the database (Table B.1). Where m is the first counted quadrat, any preceding quadrats are assigned the same urchin count as m :

$${}^{t}u_{T,q} = {}^{t}u_{T,m} \text{ for } q < m$$

applies to the deep end of transects. This does not normally occur on surveys but is implemented in the program so that it can handle such cases if they arise.

Similarly, where n is the last counted quadrat, any quadrats lacking count data that follow are assigned the same count as n :

$${}^{t}u_{T,q} = {}^{t}u_{T,n} \text{ for } q > n$$

applies to the shallow end of transects. This does not normally occur on surveys but is implemented in the program so that it can handle such cases if they arise.

For an uncounted quadrat, q, that occurs between counted quadrats j and k, linear interpolation is used to estimate the number of urchins:

$${}^{t}u_{T,q} = {}^{t}u_{T,j} + \frac{q-j}{k-j} * \left({}^{t}u_{T,k} - {}^{t}u_{T,j}\right)$$

Figure 1 visually demonstrates how linear interpolation is used to estimate urchin counts on a subsampled Red Sea Urchin transect.

2.4.2.3 Deterministic Approximation of Size Probabilities

For measured urchins in a quadrat, the size probability is determined by simply binning the animal measurements into size categories. Size frequency, ${}^{m}\vec{u}_{T,q}$, is the number of urchins within a size category in a quadrat , q, on transect, T. Size probabilities are assumed to be proportional to the size frequency of the measured urchins, and is defined as:

$$\vec{p}_{T,q} = {}^{m}\vec{u}_{T,q} / {}^{m}u_{T,q}$$

For quadrats where urchins were not measured, linear interpolation is used to estimate the size probabilities, $\vec{p}_{T,q}$. Where m is the first measured quadrat, the same size probability is assigned to any preceding quadrats:

$$\vec{p}_{T,q} = \vec{p}_{T,m} \text{ for } q < m.$$

Similarly, where n is the last measured quadrat, the same size probability is assigned to any following quadrats:

$$\vec{p}_{T,q} = \vec{p}_{T,n} \text{ for } q > n.$$

For a quadrat, q, where there are no test diameter measurements, and that occurs between measured quadrats j and k, linear interpolation is used to estimate the size probabilities:

$$\vec{p}_{T,q} = \vec{p}_{T,J} + \frac{q-j}{k-j} * (\vec{p}_{T,k} - \vec{p}_{T,j})$$

For a quadrat, q, where some urchins are measured and some are not, the size probability of the unmeasured urchins is assumed to be proportional to the size probability of the measured urchins:

$$^{u}\vec{u}_{T,q} = {}^{u}u_{T,q} * \vec{p}_{T,q}$$

This occurs rarely and is included in the programs so that they are robust to those occurrences.

2.5 ALLOMETRIC RELATIONSHIPS

Urchin weight is not measured during dive surveys. To estimate biomass, the test diameters of measured urchins must be converted to weights using allometric relationships. These weights are then used to calculate the size category-specific mean weights for each transect. These mean weights, combined with estimated number of unmeasured urchins and unmeasured urchin size proportions, are used to estimate the biomass of unmeasured urchins in the transect for each size category.

2.5.1 Red Urchin Allometric Relationship

Weight-at-length data for Red Sea Urchins were obtained from 13 surveys conducted in Alert Bay, Kendrick Island, Louise Island, Price Island and Tofino, from 1995 to 1998. All data from all samples were pooled to estimate the allometric relationship between weight and length of Red Sea Urchins. Parameter values were estimated using linear regression of the log-transforms of length and weight ($R = 0.992$; $R^2 = 0.985$). Weight (g), Ψ, for measured Red Sea Urchins with recorded test diameters (mm), Ω, is estimated with an allometric relationship:

$$\text{Ln}(\Psi_{T,q,r}) = -6.89101 + 2.7787 * \log(\Omega_{T,q,r})$$

$$\Psi_{T,q,r} = \alpha\Omega^{\beta}_{T,q,r} = 0.0010169(\Omega_{T,q,r})^{2.7787}$$

Where $\alpha = 0.0010169$ and $\beta = 2.7787$.

The estimated values for these parameters have been updated from Campbell 1998 and Leus et al. 2014 to reflect additional data that have become available. Sample size has increased from 167 (minimum 10 to maximum 150 mm test diameter) in Campbell 1998 to 5826 (minimum 6 to maximum 159 mm test diameter) here. Calculation of the Red Sea Urchin allometric relationship parameters were done outside RUAP and the program uses fixed values of α and β for all Red Sea Urchin analyses. Calculations of Red Sea Urchin weights are deterministic.

2.5.2 Green Sea Urchin Allometric Relationship

Some Green Sea Urchin surveys have survey-specific length-weight data (from dissections of urchins collected during some surveys) while other surveys do not. For surveys with survey-specific length-weight data, calculations of Green Sea Urchin allometric length-weight relationship are performed within GUAP. Length-weight dissection data are pooled from all transects within a transect category. For surveys without length-weight data, default values that have been estimated from all survey dissection data collected up to and including 2010 are used, or the user can specify new parameters to input on the user interface screen.

The overall allometric relationship includes data from 29 surveys, 15 from PFMA 12, 10 from PFMA 18, two from PFMA 19, one from PFMA 20 and one from both PFMAs 18 and 19. The number of biological samples collected per survey ranged from 17 to 286 (DFO 2014). The overall allometric relationship is based on a hierarchical model and therefore each survey is weighted equally regardless of the number of samples collected. The overall allometric relationship reflects the among-sublocation/year variability in mean weight-at-length that has been observed in Pacific coast Green Sea Urchin data collected to date.

The allometric equation used for Green Sea Urchins assumes a linear relationship between the log of test diameter and the log of urchin weight, as recorded in the database (Table B.1, Appendix B.6).

The GSU allometric relationship is:

$$\Psi_{T,q,r} = \exp\left(\log(\alpha_{30}) + \beta * \log(\frac{\Omega_{T,q,r}}{30}) + \varepsilon_{T,q,r}\right) = \alpha_{30}\left(\left(\frac{\Omega_{T,q,r}}{30}\right)^{\beta}\right)e^{\varepsilon_{T,q,r}}$$

$$\alpha_{30} \sim Norm(\mu_{\alpha}, \sigma_{\alpha}^2)$$

$$\beta \sim Norm(\mu_{\beta}, \sigma_{\beta}^2)$$

$$\varepsilon_{T,q,r} \sim N(0, \sigma^2)$$

Note that stochasticity is introduced in the calculation of Green Sea Urchin weights. When calculating deterministic estimates of biomass for Green Sea Urchins (Section 2.7) stochasticity is removed by setting equation parameters as follows:

$\alpha_{30} = \mu_\alpha$ and $\sigma_\alpha^2 = 0$

$\beta_{30} = \mu_\beta$ and $\sigma_\beta^2 = 0$

$\sigma^2 = 0$

Default parameters are as follows:

$\log(\alpha_{30}) \sim Norm(2.41753, 0.129231^2)$

$\beta \sim Norm(2.754085, 0.00699^2)$

$\varepsilon_{T,q,r} \sim N(0, 0.122298^2)$

Use of the average parameter values in deterministic calculations gives:

$$\Psi_{T,q,r} = \exp\left(2.41753 + 2.754085 * \log\left(\frac{\Omega_{T,q,r}}{30}\right) + 0\right)$$

$$\Psi_{T,q,r} = 11.2181\left(\frac{\Omega_{T,q,r}}{30}\right)^{2.754085}$$

2.5.2.1 Estimating Green Sea Urchin Allometric Parameter Values from Transect Category Data

Where test diameter and weight measurements are available for a transect category, GUAP can estimate allometric parameter values corresponding to that particular category of transects. Linear regression is used and the calculations build upon methods given in Draper and Smith 1998:

Transformed Values:

$$y_{T,q,r} = \log\left(\Psi_{T,q,r}\right)$$

$$x_{T,q,r} = \log\left(\frac{\Omega_{T,q,r}}{30}\right)$$

Sample Size:

$$n = length(x)$$

Sample Averages:

$$\bar{x} = average(x)$$

$$\bar{y} = average(y)$$

Corrected Sum of Squares on x:

$$S_{xx} = \sum\left(x_{T,q,r} - \bar{x}\right)^2$$

Corrected Sum of Cross Products on x and y:

$$S_{xy} = \sum\left(x_{T,q,r} - \bar{x}\right) * \left(y_{T,q,r} - \bar{y}\right)$$

Estimated Slope:

$$\mu_\beta = S_{xy} / S_{xx}$$

Estimated Intercept:

$$\mu_\alpha = \bar{y} - \mu_\beta * \bar{x}$$

Fitted Values:

$$\hat{y}_{T,q,r} = \mu_{\alpha} - \mu_{\beta} * x_{T,q,r}$$

Residuals:

$$\varepsilon_{T,q,r} = y_{T,q,r} - \hat{y}_{T,q,r}$$

Root Mean Square of Residuals:

$$s_{within} = \sqrt{\frac{\sum \varepsilon_{T,q,r}^{2}}{n-2}}$$

Standard Error of Estimated Slope:

$$s_{\beta} = \frac{s_{within}}{\sqrt{S_{xx}}}$$

Standard Error of Estimated Intercept:

$$s_{30} = s_{within} * \sqrt{\frac{\left(\sum x_{T,q,r}^{2}\right)}{n * S_{xx}}}$$

2.6 MEAN WEIGHT

Mean weights are estimated for each size category in each transect ($W_{T,f}$). The weight of each measured urchin is approximated using the allometric relationship. Size category averages are calculated from these estimated weights and applied to the unmeasured urchins.

2.7 DETERMINISTIC ESTIMATION OF BIOMASS

Deterministic estimates of transect biomass are used to generate estimates of mean biomass for a transect category.

Biomass of urchins on a transect comes from both measured and unmeasured urchins. The biomass of the measured urchins on a transect, ${}^{m}B_{T,f}$, is approximated by adding up the estimated weights of the all measured urchins.

The biomass of unmeasured urchins in a quadrat q on transect T is approximated from the size category mean weights for the transect multiplied by the estimated number of unmeasured urchins in size category f:

$${}^{u}b_{T,q,f} = {}^{u}u_{T,q,f} * W_{T,f}$$

The biomass of unmeasured urchins of size category f on transect T is therefore:

$${}^{u}B_{T,f} = \sum_{q} {}^{u}b_{T,q,f}$$

Total biomass for a given size category and transect (${}^{t}B_{T,f}$) is the sum of the biomass of measured and unmeasured urchins in the given size category f, for the given transect T:

$${}^{t}B_{T,f} = {}^{m}B_{T,f} + {}^{u}B_{T,f}$$

2.8 STOCHASTIC ESTIMATES OF ABUNDANCE

For measured urchins, the population, ${}^{m}u_{T,q}$, and size frequency, ${}^{m}\vec{u}_{T,q}$, are the same for stochastic calculations as they are for deterministic calculations. For red urchins, the weight of a measured urchin in a stochastic calculation is the same as the weight for a deterministic

calculation. For green urchins and stochastic calculations, the weight of a measured urchin will vary as per the stochasticity incorporated into the allometric equation.

For unmeasured urchins in the stochastic calculations, if there are multiple size categories, there is variability in the size probabilities, ${}^{*}\vec{p}_{T,q}$, which propagates into variability in the size frequencies, ${}^{*\mu}\vec{u}_{T,q}$, and ultimately into variability in the estimates of population and biomass density.

For a measured quadrat, the size probability is sampled according to the dirichlet distribution:

$${}^{*}\vec{p}_{T,q} \sim dirichlet({}^{m}\vec{u}_{T,q})$$

The size probabilities for the measured quadrats are then interpolated to unmeasured quadrats. Interpolations are done as they were for the deterministic calculations. The result is a stochastic size probability for every quadrat in the transect.

2.8.1 Stochastic Estimates of Number of Urchins

For unmeasured urchins in a quadrat q on transect T, the stochastic estimate of the number of urchins, in each size category, are sampled according to the multinomial distribution:

$${}^{*\mu}\vec{u}_{T,q} \sim multinomial({}^{u}u_{T,q}, {}^{*}\vec{p}_{T,q})$$

Total stochastic estimate of the number of urchins in a given size category and transect (${}^{*\iota}\vec{u}_{T}$), is the sum of the number of measured and unmeasured urchins in the given size category.

$${}^{*\iota}\vec{u}_{T} = {}^{m}\vec{u}_{T} + {}^{*\mu}\vec{u}_{T}$$

${}^{*\iota}\vec{u}_{T}$ is then eligible to use as a measure of abundance with the ratio estimator (see Section 2.9)

2.8.2 Stochastic Estimates of Biomass

The stochastic estimate of biomass for unmeasured urchins of size category f, quadrat q on transect T is:

$$^{*u}b_{T,q,f} = {}^{*u}u_{T,q,f} * W_{T,f}$$

The stochastic estimate of biomass of unmeasured urchins of size category f on transect T is therefore:

$$^{*u}B_{T,f} = \sum_{q} {}^{*u}b_{T,q,f}$$

Total stochastic estimate of biomass for a given size category and transect ($^{*t}\bar{B}_T$) is the sum of the biomass of measured and unmeasured urchins in the given size category.

$$^{*t}\bar{B}_T = {}^{m}\bar{B}_T + {}^{*u}\bar{B}_T$$

$^{*t}\bar{B}_T$ is then eligible to use as a measure of abundance with the ratio estimator (see Section 2.9)

In summary, each time a stochastic estimate of abundance is generated for a transect, it is a matter of:

- resampling size probabilities for the measured quadrats,
- interpolating size probabilities for the unmeasured quadrats
- resampling number of urchin in each size category for unmeasured urchins
- calculating mean weight of urchins in each size category on the transect
- calculating biomass for the unmeasured urchins
- adding the estimated abundance (population and biomass) of the unmeasured urchins to the measured urchins

2.9 THE RATIO ESTIMATOR

For a category of transects and for each specified size category (*f*) of urchins, mean abundance, D_f, is estimated with confidence bounds using a ratio estimator (Cochran 1977) and bootstrapping (Efron and Tibshirani 1993).

Density is estimated using deterministic extrapolation and, using the basic form of the ratio estimator, is:

$$D_f = \frac{\sum_T A_{T,f}}{\sum_T S_T}$$

Where $A_{T,f}$ = estimated abundance of urchins (population ${}^{*J}\bar{n}_T$ or biomass ${}^{*J}\bar{B}_T$) of size category *f*, on transect *T*, and S_T is the size of transect *T*.

The calculation requires an abundance, $A_{T,f}$, and a size, S_T, for each transect.

In most applications of the ratio estimator, $A_{T,f}$ and S_T are exact values and therefore D is deterministic. Deterministic estimates of density are made because it is believed that these are the most probable values and because these most probable results are required as inputs to estimate confidence bounds (see Bootstrapping).

RUAP and GUAP also incorporate ${}^{*}A_{T,f}$, a stochastic value of transect abundance, based on stochastic extrapolation, which is used to generate ${}^{*}D$, a stochastic estimate of density as described earlier. Estimates of ${}^{*}D$ are used in the bootstrapping process (section 2.9.1) to produce confidence bounds.

Transect size, S_T, is measured in two ways: for linear calculations, the size of a transect is measured by its width, whereas for spatial calculations, transect size is the area of sea floor (square metres) within the specified depth range. Transect size is simplified by the fact that all urchin transects are one metre wide. For linear calculations, S_T equals one metre and the ratio estimator reduces to a simple average. For spatial calculations, the transect size is the length or area of the transect which is in the specified depth range, defined as the count of quadrats (surveyed and unsurveyed) in the depth range.

2.9.1 Bootstrapping

Bootstrapping (Efron and Tibshirani 1993) is a resampling technique used to estimate the accuracy of values such as the ratio estimator, which in this case is an estimate of density. For each transect category, transects are resampled with replacement. The number of resampled transects equals the number that originally occurred in the transect category. Each time the resampling occurs, it results in a new set of transects, τ_i. In τ_i, some transects can occur several times where others may not occur at all. The ratio estimator, ${}^*D_i$ is calculated for each occurrence of τ_i. The default number of resamples is 10,000 ($i = 1,2,3,...10,000$).

In a standard application of bootstrapping, ${}^*A_{T,f}$ would be an exact value and the distribution of ${}^*D_i$ would reflect uncertainty due to transect-to-transect variability. In the case of urchin surveys, unsurveyed quadrats result in uncertainty of transect abundance which contributes to the uncertainty of mean abundance. This extra uncertainty is incorporated by generating a new value of ${}^*A_{T,f}$ every time a transect is sampled in the bootstrapping process.

$$ {}^*D_{i,f} = \frac{\sum_{T \in \tau_i} {}^*A_{T,f}}{\sum_{T \in \tau_i} S_T} $$

The frequency distribution of ${}^*D_{i,f}$ reflects uncertainty in the transect abundances as well as transect-to-transect variability.

Confidence bounds on D are generated from a sample of ${}^*D_{i,f}$ values. The distribution of ${}^*D_{i,f}$ is used as an approximation to the probability density function for D. Bias-corrected accelerated percentile intervals (BC_a) (Efron and Tibshirani 1993) are used to calculate confidence intervals on D. BC_a calculations require the estimated deterministic value of mean density (D) as well as the sample of ${}^*D$ values.

2.9.2 Random Seeds

The confidence bounds generated by bootstrapping in GUAP and RUAP are calculated using random values. There is a single implementation of a random seed each time GUAP or RUAP is

run. As long as the same random seed is used and calculations are applied to identical datasets, the confidence bounds are repeatable.

2.10 OUTPUTS

Output from RUAP and GUAP is in the form of a .mdb file. In order to achieve a higher degree of compatibility for the output, an older version of OLE DB (4.0) is used to write the output. OLD DB (4.0) was released with Microsoft ACCESS 2000.

The output consists of related tables that report mean abundance (density and biomass) with confidence bounds, linearly and spatially, for each combination of size and transect category, and metadata (analysis date, number of iterations, random seed, minimum depth, maximum depth).

3 DISCUSSION

RUAP and GUAP are part of a family of computer programs used primarily for routine analysis of survey data collected by the Pacific Region's dive programs at DFO. The other two computer programs in the family are used to analyze Pacific Geoduck (*Panopea generosa*) and Giant Red Sea Cucumber (*Parastichopus californicus*) data. There are efficiencies in developing and maintaining these computer programs as a family; ultimately there is less code to develop and maintain. The same abstractions developed to simultaneously support multiple dive programs will be useful for addressing the evolving needs of these programs.

GUAP and RUAP read survey data from its final destination, a Microsoft SQL Server biological database maintained by DFO's Shellfish Data Unit, or a database of the same structure. This streamlines the workflow by eliminating any manipulation of data that would otherwise have to occur prior to analysis. The other advantage is that it reinforces an orderly data-flow – data has to be error-free and in its final destination before it can be analyzed.

It is expected that PYTHON 3 will have longevity as a suitable platform for the analysis programs. There is an ever increasing amount of open source, statistical and mathematical software available in PYTHON (scipy.org for example). PYTHON also has many uses beyond the statistical and mathematical world (see PYTHON.org for examples).

The first component of the analysis programs to become obsolete will likely be the file reading and writing libraries (OLE DB). Microsoft OLE DB and ACCESS are declining in popularity. If the biological databases are ported to a different platform, the analysis programs are well modularized in that respect and therefore should easily be modified to use the new data-source.

What could be more problematic is if .mdb or .accdb files cease to be a practical format for the output of the analysis programs. Most other databases are server-based as opposed to file-based (like OLE DB) and are impractical to use with small, temporary sets of data. Major effort would be required to respond to such a scenario.

The aspect of RUAP and GUAP that could be considered novel is the way linear interpolation is used to estimate data for unsurveyed quadrats. The approach is based upon the assumption that unsurveyed quadrats are similar to the nearest surveyed quadrats. It provides flexibility in the development of a survey plan that matches the available resources to do the fieldwork. The linear interpolation approach is also robust with respect to deviations from a survey plan, which occurred more frequently in older surveys but can still occur due to unanticipated logistical and safety issues in more recent surveys.

4 ACKNOWLEDGEMENTS

Many thanks to Dominique Bureau, Nicholas Duprey, Claudia Hand and Joanne Lessard for their comments and suggestions during the preparation of this document.

5 REFERENCES

Campbell, A. 1998. Catch, Effort and Quota Estimates for the Red Sea Urchin Fishery in British Columbia, p.83-109. In: B. J. Waddell, G. E. Gillespie, and L. C. Walthers [eds.]. Invertebrate Working Papers reviewed by the Pacific Stock Assessment Review Committee (PSARC) in 1995. Part 2. Echinoderms. Can. Tech. Rep. Fish. Aquat. Sci. 2215.

Cochran, W.G. 1977. Sampling Techniques. Chapman & Hall, New York, New York.

DFO. 2014. Stock Status Update and Quota Options for the Green Sea Urchin (*Strongylocentrotus droebachiensis*) Fishery in British Columbia, 2013-2016. DFO Can. Sci. Advis. Sec. Sci. Resp. 2014/052.

Draper, N.R. and Smith, H. 1998. Applied Regression Analysis. John Wiley and Sons, New York, New York.

Efron, B. and Tibshirani, R.J. 1993. An Introduction to the Bootstrap. Chapman & Hall, New York, New York.

Leus, D., Campbell, A., Merner, E., Hajas, W.C., and Barton, L.L. 2014. Framework for Estimating Quota Options for the Red Sea Urchin (*Strongylocentrotus franciscanus*) Fishery in British Columbia Using Shoreline Length and Linear Density Estimates. DFO Can. Sci. Advis. Sec. Res. Doc. 2013/094. vi + 69 p.

Microsoft. OLE DB. Microsoft Corporation. Redmond, Washington.

Waddell, B.J. and Perry, R.I. 2012. Survey Results of Green Sea Urchin (*Strongylocentrotus droebachiensis*) Populations in Queen Charlotte Strait, British Columbia, October 2008 and November 2010. Can. Tech. Rep. Fish. Aquat. Sci. 3000: ix + 73 p.

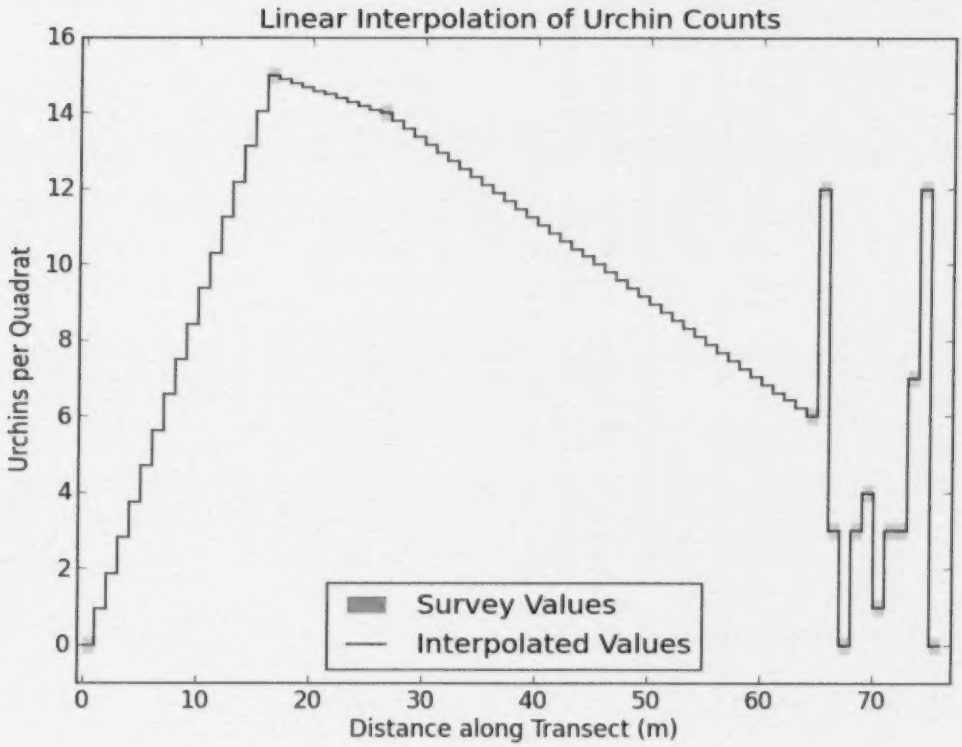

Figure 1. Linear interpolation as used to approximate urchin counts from a subsampled Red Sea Urchin transect.

6 APPENDIX A. DESCRIPTION OF VARIABLES

VARIABLE	DESCRIPTION
${}^{m}b_{T,q}$, ${}^{u}b_{T,q}$ and ${}^{t}b_{T,q}$	The biomass of measured, unmeasured and total urchins in the q^{th} quadrat of the T^{th} transect
${}^{m}\vec{u}_{T,q} = \left({}^{m}u_{T,q,1}, {}^{m}u_{T,q,2}, \ldots, {}^{m}u_{T,q,F}\right)$	The number of measured urchins in the q^{th} quadrat of the T^{th} transect. One value for each size category.
$\vec{p}_{T,q} = (p_{T,q,1}, p_{T,q,2}, \ldots, p_{T,q,F})$	The probability associated with each size category. Specific to the q^{th} quadrat of the T^{th} transect.
$\vec{W}_T = (W_{T,1}, W_{T,2}, \ldots, W_{T,F})$	The estimated mean weight for each size category. Specific to the T^{th} transect.
$A_{T,f}$	The abundance of urchins in a given size category f and in the T^{th} transect. (deterministic)
S_T	The size of the T^{th} transect
D_f	Estimated mean abundance for a given size category and category of transects (deterministic).
$\Omega_{T,q,r}$	The test diameter of the r^{th} measured urchin from the q^{th} quadrat of the T^{th} transect.
$\Psi_{T,q,r}$	The estimated weight of the r^{th} measured urchin from the q^{th} quadrat of the T^{th} transect.
${}^{t}u_{T,m}$, ${}^{t}u_{T,q}$ and ${}^{t}u_{T,n}$	The total number of urchins in the m^{th} , q^{th} and n^{th} quadrat of the T^{th} transect.

${}^{m}\vec{u}_{T,q} = \left({}^{m}u_{T,q,1}, {}^{m}u_{T,q,2}, \ldots, {}^{m}u_{T,q,F}\right)$	The number of urchins in each size category (1 to F) of the measured urchins in the q^{th} quadrat of the T^{th} transect.
${}^{m}u_{T,q}$	The number of measured urchins in the q^{th} quadrat of the T^{th} transect.
$\vec{p}_{T,m} = \left(p_{T,m,1}, p_{T,m,2}, \ldots, p_{T,m,F}\right)$, $\vec{p}_{T,q} = \left(p_{T,q,1}, p_{T,q,2}, \ldots, p_{T,q,F}\right)$ and $\vec{p}_{T,n} = \left(p_{T,n,1}, p_{T,n,2}, \ldots, p_{T,n,F}\right)$	The size probabilities for the the m^{th} , q^{th} and n^{th} quadrat of the T^{th} transect.
${}^{*}A_{T,f}$	The stochastic estimate of abundance of urchins in T^{th} transect, in size category f.
${}^{*}D_{i}$	The i^{th} stochastic estimate of mean density for a transect category.
${}^{*}\tau_{i}$	The i^{th} randomly sampled (with replacement) set of transects taken from the current transect category
α and β	Coefficients used to estimate the weight of a Red Sea Urchin given its test diameter.
$\varepsilon_{T,q,r}$	The error term that occurs when the test diameter is used to estimate the weight of the r^{th} measured Green Sea Urchin of the q^{th} quadrat of the T^{th} transect.
α_{30} and β	Coefficients used to estimate the weight of a Green Sea Urchin given its test diameter.

${}^{u}b_{T,q,f}$	The deterministic estimate of biomass for unmeasured urchins of the f^{th} size category in the q^{th} quadrat of the T^{th} transect.
${}^{u}u_{T,q,f}$	The number of unmeasured urchins of the f^{th} size category in the q^{th} quadrat of the T^{th} transect.
$W_{T,f}$	Mean weight of the f^{th} size category of urchins in the T^{th} transect.
${}^{u}B_{T,f}$	Biomass of unmeasured urchins in the f^{th} size category of urchins in the T^{th} transect.
${}^{*,u}u_{T,q} = ({}^{*,u}u_{T,q,1}, {}^{*,u}u_{T,q,2}, \ldots, {}^{*,u}u_{T,q,F})$	A stochastic estimate of the number of unmeasured urchins in the q^{th} quadrat of the T^{th} transect.
${}^{*,u}b_{T,q,f}$	The stochastic estimate of biomass for unmeasured urchins of the f^{th} size category in the q^{th} quadrat of the T^{th} transect.
$d_{T,q}$	Depth of the q^{th} quadrat of the T^{th} transect.

7 APPENDIX B. SOFTWARE STRUCTURE

RUAP and GUAP were developed in PYTHON 3.3 (van Rossum, 1995). Further mathematical functionality is achieved through the NumPy and SciPy libraries (Oliphant 2007). Python is object based (Rumbaugh et al. 1991). The two programs share much of the same code. The code used by the software can be viewed on-line at the following Uniform Resource Locator (url):

https://drive.google.com/folderview?id=0B-07f3-Z5WQKRnpIbmdZMXpMYkU&usp=sharing

In this document, except where specifically noted, the same calculations are being used in both RUAP and GUAP.

The Red and Green Sea Urchin biological database relationships are illustrated in Figure B.1. and B.2.

7.1 APPENDIX B.1. THE UNMEASANIMALS CLASS

The UnMeasAnimals class is used to represent a group of animals for which sizes are not available.

7.1.1 Input to the constructor:

- The number of unmeasured animals
- Size bounds
- Average weight per size category for a given transect ($W_{T,f}$)

7.1.2 Methods of Note:

- Set the size probability ($\vec{p}_{T,q}$)
- Set the average weights ($W_{T,f}$)
- Get abundance($^{u}u_{T,q,f}$ and $^{u}b_{T,q,f}$) for each size category

7.1.3 Other Notes:

- The number of unmeasured animals can be zero – in which case many features of the class are trivial.

7.2 APPENDIX B.2. THE MEASANIMALS CLASS

MeasAnimals is used to represent a group of animals for which there are size measurements.

7.2.1 Input to the constructor:

- Sizes of the animals ($\Omega_{T,q,r}$)
- Size bounds

7.2.2 Methods of Note:

- Return a random or deterministic size probability corresponding to the size frequency, ${}^{m}\vec{u}_{T,q}$, of the measured animals.
- Return population and estimated biomass for all size categories (${}^{m}u_{T,q,f}$ and ${}^{m}b_{T,q,f}$). As input, a function for the allometric relationship is required.

7.2.3 Other Notes:

- The number of measured animals can be zero – in which case many features of the class are trivial.

7.3 APPENDIX B.3. THE QUADRAT CLASS

Quadrat is used to represent a quadrat, surveyed or unsurveyed. It is an aggregate of MeasAnimals and UnMeasAnimals

7.3.1 Input to the constructor:

- Quadrat number
- Depth ($d_{T,q}$)
- Number of animals(${}^{t}u_{T,q}$)
- Size measurements for the quadrat ($\Omega_{T,q,r}$)
- Size bounds

7.3.2 Methods of Note:

- Randomize the size-probability.(${}^{*}\vec{p}_{T,q}$)
- Get abundance of measured animals.(${}^{m}u_{T,q,f}$ and ${}^{m}b_{T,q,f}$)
- Get abundance of unmeasured animals.(${}^{u}u_{T,q,f}$ and ${}^{m}b_{T,q,f}$)

7.4 APPENDIX B.4 THE TRANSECT CLASS

Transect is used to represent a survey-transect; from the first surveyed quadrat to the last surveyed quadrat. Transect is an aggregate of quadrat-objects.

7.4.1 Input to the constructor:

- A connection to a GSU or RSU biodatabase
- A transect-key from the biodatabase
- A source of the allometric equation
- The species specific library (see Appendices B.8 and B.9)
- Size Bounds
- Depth Bounds

7.4.2 Methods of Note:

- Randomize the size frequency of unmeasured animals for each quadrat (${}^{*,\mu}\vec{u}_{T,q}$ values)
- Get abundance within the depth range (A_T or ${}^{*}A_T$)
- Get area within the depth range (S_T)
- Get width of transect (S_T)
- Calculate average weight of measured animals for each size category ($W_{T,f}$)

7.4.3 Other Notes:

- It is through this that unsurveyed quadrat values are estimated by linear interpolation (Section 2.4.2).

7.5 APPENDIX B.5 THE TRANSECTCATEGORY CLASS

TransectCategory is used to represent a collection of transects which in turn are assumed to be a random sample from a particular time and location. TransectCategory is an aggregate of Transect-objects.

7.5.1 Input to the constructor:

- A connection to a GSU or RSU biodatabase
- A list of transect-keys from the biodatabase
- A source of the allometric equation
- The species specific library (see Appendices B.8 and B.9)
- Size Bounds
- Depth Bounds

7.5.2 Methods of Note:

- Get average abundance.(D or $^{*}D$)
- Generate a randomized copy of itself; suitable for bootstrapping.(Section 2.9.1)
- Get confidence bounds on mean-density based upon bootstrapping.(Section 2.9.1)

7.6 APPENDIX B.6 THE DATAALLOMETRIC CLASS

DataAllometric class is used for GSU but not RSU. It is a source of allometric functions corresponding to size and weight data collected from a specific set of transects. Coefficients and associated standard errors for the allometric equation are calculated as per Section 2.5.2. The allometric function can either be based on estimated values of the coefficients or values that have been randomized according to the standard errors.

7.6.1 Input to the constructor:

- A connection to a GSU biological database
- A list of transect-keys from the biological database

7.6.2 Methods of Note:

- Get a source of allometric functions (deterministic and stochastic) corresponding to length-weight data from the specified transects.

7.6.3 Other Notes:

- SQL command to get length-weight data:

```
SELECT Dissection.Diameter, Dissection.WetWgt
FROM Dissection
WHERE(
(Dissection.Diameter Is Not Null )
and (Dissection.Diameter <>0 )
and (Dissection.WetWgt Is Not Null )
and (Dissection.WetWgt <>0 )
and(
        (Dissection.HKey=key#1) or
        (Dissection.HKey=key#2) or
        (Dissection.HKey=key#3) or
        ...
)       );
```

7.7 APPENDIX B.7 THE GSU LIBRARY

It is through this library that specific characteristics of GSU, protocols used to survey GSU and the GSU database are incorportated into TransectCategory and its constituents. In order to maintain compatibility, all libraries used this way have to have a similar structure.

7.7.1 Data Contained in the library:

- Area of a quadrat equals one metre-squared
- Width of a transect equals one metre
- Confidence levels to appear in the output are 99,95,90,75 and 50%.

7.7.2 Functions to generate SQL commands:

- Quadrat numbers, depth and counts as recorded in for a specific transect.

```
SELECT Density.QuadratNum as QuadNum
    iif (Density.GageDepth is NULL,-100,(Density.GageDepth-Density.TideHgt) *0.3048)  as Depth,
    iif (Density.CountTotal is NULL,   0,Density.CountTotal) as CountTotal
FROM Headers INNER JOIN Density ON Headers.Key = Density.HKey
WHERE (
  (Headers.Key= key# )
  AND ((Density.QuadratNum)>0) and  (Density.QuadratNum  Is Not Null)  )
ORDER BY Density.QuadratNum;
```

- Quadrat numbers where there are length measurements are available.

```
Select Distinct SF.QuadratNum as QuadNum
FROM Density INNER JOIN SF ON Density.HKey = SF.HKey
WHERE (Density.HKey=key#)
order by SF.QuadratNum;
```

- Test diameters from a specific quadrat.

```
SELECT  SF.Diameter as AnimalLength
```

```
FROM Density INNER JOIN SF ON (Density.QuadratNum = SF.QuadratNum)
        AND (Density.HKey = SF.HKey)
WHERE (
        (Density.HKey= key#)
        AND (Density.QuadratNum= quad#)
        And (SF.Diameter Is Not Null)
        AND ((SF.Sp)='G' Or (SF.Sp)='' Or (SF.Sp) Is Null));
```

7.8 APPENDIX B.8 THE RSU LIBRARY

It is through this library that specific characteristics of RSU, protocols used to survey RSU and the RSU database are incorportated into TransectCategory and its constituents. In order to maintain compatibility, all libraries used this way have to have a similar structure.

7.8.1 Data Contained in the library:

- Area of a quadrat equals one metre-squared
- Width of a transect equals one metre
- Confidence levels to appear in the output are 99,95,90,75 and 50%.

7.8.2 Functions to generate SQL- commands:

- Quadrat numbers, depth and counts as recorded in for a specific transect.

```
SELECT Density.QuadratNum as QuadNum,
  iif (Density.ChartDepth is NULL,-100,Density.ChartDepth)as  Depth,
  iif (Density.CountTotal is NULL,  0,Density.CountTotal) as CountTotal
FROM Headers INNER JOIN Density ON Headers.Key = Density.HKey
 WHERE (
  (Headers.Key= key# )
  AND ((Density.QuadratNum)>0) and  (Density.QuadratNum  Is Not Null)  )
 ORDER BY Density.QuadratNum;
```

- Quadrat numbers where there are length measurements available.

```
Select Distinct SF.QuadratNum as QuadNum
 FROM Density INNER JOIN SF ON Density.HKey = SF.HKey
WHERE (
        (Density.HKey=key#)
        and (Density.CountTotal>0) and (SF.QuadratNum is not Null) and
         ((SF.Sp='R') Or (SF.Sp='') Or (SF.Sp Is Null)) )
```

```
order by SF.QuadratNum;
```

- Test diameters from a specific quadrat.

```
SELECT  SF.Diameter as AnimalLength
FROM Density INNER JOIN SF ON (Density.QuadratNum = SF.QuadratNum)
        AND (Density.HKey = SF.HKey)
WHERE (
        (Density.HKey = key#)
        AND (Density.QuadratNum = quad#)
        And (SF.Diameter Is Not Null) AND
          ( (SF.Sp ='R') Or (SF.S) ='')  Or (SF.Sp Is Null)));
```

7.9 APPENDIX B. REFERENCES

Oliphant,T.E. 2007. Python for Scientific Computing. Computing in Science and Engineering. 9,3:10-20.

Rumbaugh, J., Blaha M., Premerlani, W., Eddy, F. and Lorenden,W. 1991. Object-Oriented Modeling and Design. Prentice Hall, Engle Wood Cliffs, New Jersey.

van Rossum, G. 1995. Python Tutorial. Technical Report CS-R9526. Centrum voor Wiskunde en Informatica (CWI), Amsterdam.

Table B.1. Data from the Red Sea Urchin and Green Sea Urchin biological databases used by RUAP and GUAP to estimate mean density. Key and Hkey are used to construct joins between tables within the database. QuadratNum is used to construct joins of quadrats within a transect.

Table or Query Name	Field Name	Data Type	Symbol in Current Document (where applicable)	Can be Used to Define Transect-Category
Headers	InBed (RUAP only)	Byte		yes
	Key	Long Integer		
	SurveyTitle	Text(Field Size=50)		yes
	Location (RUAP only)	Text(Field Size=50)		yes
	SubSampleLocation (GUAP only)	Text(Field Size=50)		yes
	SiteNum (RUAP only)	Byte		yes
	StatArea	Byte		yes
	SubArea	Byte		yes
	TransectNum	Text		yes
	Year	Integer		yes
Density	Hkey	Long Integer		
	CountTotal	Integer	$'u_{T,q}$	
	QuadratNum	Integer		
	GageDepth(GUAP only)	Byte		
	TideHgt(GUAP only)	Real		
	ChartDepth	Single	$d_{T,q}$	
SF	Hkey	Long Integer		
	QuadratNum	Integer		
	Diameter	Integer	$\Omega_{T,q,r}$	
	Sp	Text(Field Size=255)		
Dissection (GUAP only)	HKey	Long Integer		
	Diameter	Integer	$\Omega_{T,q,r}$	
	WetWgt	Single	$\Psi_{T,q,r}$	

Figure B.1. Red Sea Urchin biological database relationships

Figure B.2. Green Sea Urchin biological database relationships